

STANDARDIZACE METADAT PRO NÁRODNÍ REGISTR EVŠKP

Eva Bratková a Jan Mach

Bibliografický záznam

BRATKOVÁ, E.; MACH, J. Standardizace metadat pro národní registr EVŠKP. In MIKULECKÁ, J.; SEDLÁČEK, J. (ed.). *Digitální knihovny : provoz a budování : sborník konference EUNIS-CZ, Špindlerův Mlýn, 28. až 30. května 2006*. Vyd. 1. s přílohou CD. Hradec Králové : Gaudeamus, 2006, s. 5-12. ISBN 80-7041-200-3.

Abstrakt

Příspěvek je věnován představení návrhu národního Metadatového souboru pro elektronické vysokoškolské kvalifikační práce (EVSKP-MS, verze 0.1), připraveného v rámci aktivit vysokoškolských knihoven. V úvodní části jsou charakterizovány cíle, význam a východiska jeho vytvoření. Druhá část zahrnuje stručný popis specifikace navrhovaných metadatových prvků, zdůrazněny jsou prvky povinné pro potřeby jejich komunikace na národní a mezinárodní úrovni. Bližší komentář se týká zejména problémových prvků. Ve třetí části jsou uvedeny v návrhu doporučené možnosti zápisu metadat ve formátu RDF/XML, XML a HTML/XHTML a zápis obsahu některých prvků – data, jazyka, formátu zdroje a jména fyzických osob.

Klíčová slova

Elektronické vysokoškolské kvalifikační práce, Metadata, Metadatové formáty, Dublin Core

1 Úvod

Jedním z úkolů, který začala Odborná komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací AKVŠ (<http://www.evskp.cz/>) ihned po svém založení na jaře 2004 řešit, bylo vytvoření návrhu **národního metadatového standardu pro popis elektronických vysokoškolských kvalifikačních prací (eVŠKP)**. Důvodem byla skutečnost, že v té době již některé vysoké školy v ČR zahájily provoz nebo připravovaly projekt lokálního systému pro zpřístupňování eVŠKP, v jehož rámci byly připraveny buď vlastní lokálně platné specifikace metadat, nebo byly adaptovány specifikace z jiných systémů (domácích nebo zahraničních). Velká variabilita lokálních specifikací je velkou překážkou pro budování plánovaného **národního registru eVŠKP** a popřípadě i **národního elektronického archivu úplných textů eVŠKP**, a proto bylo rozhodnuto připravit návrh tohoto standardu co nejdříve. Záměrem bylo, aby se národní standard stal pro lokální specifikace závazný, protože jen tak bude možná bezproblémová směna metadat mezi systémy a bude možná zároveň i integrace metadat o českých eVŠKP do mezinárodních systémů.

Přípravou standardu byla pověřena Pracovní skupina pro metadata. Oficiální práce na metadatovém standardu byly zahájeny v dubnu 2004. Základním východiskem návrhu standardu se staly výsledky podrobné analýzy významných zahraničních metadatových specifikací, která proběhla již v roce 2003 v rámci průzkumu tehdejšího stavu zahraničních systémů pro zpřístupňování e-VŠKP (součást jedné z grantových úloh FRVŠ [1, část 3.2.3]). Výsledek analýzy vedl, kromě jiných faktorů, k rozhodnutí, že zpracování českého národního standardu bude v obecné rovině vázáno na specifikaci známého *Souboru metadatových prvků Dublin Core* [2] a že v rovině specifické se

v zájmu interoperability bude přihlížet ke specifikaci amerického souboru metadat pro elektronické disertační a diplomové práce *ETD-MS* [4], který využívají americké i další univerzity sdružené v mezinárodním systému NDLTD, a dále, zejména pokud jde o kvalifikující informace, ke specifikaci velmi propracovaného německého národního standardu pro vysokoškolské práce *METADISS* [6], který je společným formátem Německé knihovny a národního projektu SRN „Dissertationen Online“. Při tvorbě návrhu českého standardu sehrály důležitou roli také domácí zkušenosti s popisem těchto dokumentů ve vysokoškolských knihovnách.

V období od dubna 2004 do září 2005 vzniklo několik interních verzí návrhu standardu, které byly podrobeny připomínkování všech členů celé Odborné komise. Návrh *Metadatového souboru pro elektronické vysokoškolské kvalifikační práce (EVSKP-MS)* byl ve své verzi 0.1 zveřejněn na webových stránkách Odborné komise 7. listopadu 2005 [5]. Zainteresované komunity byly vyzvány k veřejnému připomínkování. Po vyhodnocení všech připomínek bude připravena definitivní 1. verze tohoto standardu. Návrh zahrnuje ucelenou **specifikaci jednotlivých metadatových prvků** pro popis eVŠKP obhájených na vysokých školách v ČR (viz dále kap. 2) a základní vymezení metod **reprezentace zařazených prvků** pro potřeby jejich komunikace a sdílení v prostředí počítačových sítí (viz dále kap. 3).

2 Specifikace metadat

Návrh národního metadatového souboru *EVSKP-MS* zahrnuje definice celkem **22 prvků**. Povinné jádro tvoří polovina z nich, tj. 11 prvků. V následujícím přehledu jsou povinné prvky uvedeny s podtržením, nepovinné jsou uvedeny kurzivou a v kulaté závorce je uvedena opakovatelnost prvku:

1. Název VŠKP (Neopakovatelný)
2. *Podnázev VŠKP* (Opakovatelný)
3. *Překlad názvu VŠKP* (Opakovatelný)
4. *Překlad podnázvu VŠKP* (Opakovatelný)
5. Autor VŠKP (Neopakovatelný)
6. *Datum narození autora VŠKP* (Neopakovatelný)
7. *Předmětový popis VŠKP* (Opakovatelný)
8. Abstrakt VŠKP (Opakovatelný)
9. *Instituce archivující a zpřístupňující VŠKP* (Opakovatelný)
10. *Osoby podílející se na vedení VŠKP* (Opakovatelný)
11. *Osoby podílející se na oponování VŠKP* (Opakovatelný)
12. Datum vytvoření / odevzdání VŠKP (Neopakovatelný)
13. *Datum obhájení VŠKP* (Neopakovatelný)
14. Typ VŠKP (Opakovatelný)
15. Formát VŠKP (Opakovatelný)
16. Identifikátor VŠKP (Opakovatelný)
17. Jazyk VŠKP (Opakovatelný)
18. *Práva k využívání VŠKP* (Opakovatelný)
19. Přidělovaný akademický titul (Neopakovatelný)
20. *Typ studijního programu* (Neopakovatelný)
21. Studijní program a studijní obor (Neopakovatelný)
22. Instituce přidělující titul (Neopakovatelný)

Jednotlivé prvky jsou ve specifikaci [5, kap. 2.1] uvedeny českým pojmenováním (v titulku), v tabulkovém přehledu je pak postupně prezentováno jejich označení a definované kvalifikátory. Následuje informace o opakovatelnosti prvku v jednom záznamu a informace o jeho závaznosti (povinnosti). Popis zahrnuje detailní informaci o prvku včetně jeho zápisu. Připojené příklady syntaxe jsou prozatímním doplňkem specifikace (podrobnější informace ke struktuře metadat - viz kap. 3).

K bezproblémovým prvkům, na kterých se členové Komise v zásadě shodli, patří základní popisné údaje: **Název** (povinný) a **Podnázev VŠKP**, **Překlad názvu** a **Překlad podnázvu**, **Autor** (povinný) a **Datum narození autora** (minimálně rok). Zvažována byla povinnost překladů názvů prací do cizích jazyků, zejména angličtiny, na národní úrovni. Prvky zůstávají nepovinné, ale jednoznačně je doporučen překlad názvu do angličtiny u disertačních prací. Perspektivně se počítá s propojováním jmen autorů na soubory autoritních záznamů.

V zásadě k bezproblémovým prvkům patří také **Abstrakt VŠKP** (povinný). Jeho uvádění v češtině bylo ustanoveno jednomyslně. Překlad abstraktu do angličtiny je doporučen jako povinný pouze u disertačních prací, u dalších typů VŠKP (bakalářských, diplomových nebo rigorózních) lze tuto povinnost stanovit na úrovni lokálních systémů v rámci vnitřních předpisů. Zřejmě největším problémem při tvorbě specifikace byl předmětový popis. V nejbližší budoucnosti nelze v ČR počítat s jednotným předmětovým pořadacím systémem pro potřeby eVŠKP. V rámci prvku **Předmětový popis VŠKP** je proto definován kvalifikátor, který bude specifikovat užitý pořadací systém. Předmětový popis je na národní úrovni nepovinnou záležitostí. Překlady předmětových termínů do cizích jazyků jsou jen doporučením pro lokální systémy.

Jako bezproblémové se jeví prvky určené pro zápis jmen vedoucích a oponentů VŠKP, jejich uvádění je však nepovinné, na některých pracovištích mohou nastat objektivní potíže s jejich získáváním. Jednoznačně byla přijata definice prvku **Instituce archivující a zpřístupňující VŠKP**. Složitě, a to i na mezinárodní úrovni, je definování prvku **Data vytvoření / Odevzdání VŠKP** (povinný) a **Data obhájení VŠKP**. Charakter data je nezbytné řešit pomocí kvalifikátorů. Problémem je také jejich obtížné získávání na některých školách. V další verzi metadatového standardu se u datací mohou vyskytnout změny.

V rámci standardu byla dále definována skupina 4 povinných formálních prvků: **Typ**, **Formát**, **Identifikátor** a **Jazyk VŠKP**. U většiny z nich se předpokládá využívání číselníků. U identifikátorů se prozatím předpokládá využívání adres URL, intenzivně se však v zahraničí i u nás pracuje na zavedení trvalých identifikátorů (URN aj.). Použitá URL adresa do lokálního systému by měla být pokud možno trvalá (perzistentní). Komise v dalších verzích standardu zváží formalizaci prvku **Práva k využívání VŠKP**, a to i v souvislosti s novelou Autorského zákona.

Navrhovaný metadatový standard zahrnuje také 4 specifické prvky, které nejsou obsaženy v obecné metadatové specifikaci Dublin Core: **Přidělovaný akademický titul** (povinný), **Typ studijního programu**, **Studijní program** a **studijní obor** (povinný) a **Instituce přidávající titul** (povinný). Prvky byly adaptovány z amerického souboru *ETD-MS*. Komise zvážila jejich význam pro potřeby online vyhledávání a stanovila

většinu z nich jako prvky povinné, přestože se v současné době u některých jeví jako problém získávání příslušných hodnot. U většiny se předpokládá využívání číselníků.

3 Reprezentace metadatových prvků

Součástí návrhu národního metadatového souboru *EVSKP-MS* jsou příklady syntaxe zápisu metadat, které by měly být návodem pro správce jednotlivých lokálních systémů, jak zajistit export dat do připravovaného národního registru. Při návrhu se Komise snažila neomezovat jednotlivé lokální systémy definicí jen jednoho povinného standardu zápisu a tak podobně jako standard Dublin Core [3] nabízí hned několik možných syntaxí zápisu. Jedná se konkrétně o následující syntaxe:

- RDF/XML
- XML
- HTML/XHTML

Národní registr metadat bude podporovat při importu všechna výše uvedená doporučení a umožní vzájemný převod mezi jednotlivými způsoby reprezentace metadat. Interní reprezentace metadat v lokálních systémech není nijak předepisována a je plně na tvůrcích každého jednotlivého systému.

Detailní popis kódování metadat je popsán v návrhu národního metadatového souboru *EVSKP-MS* [5, kap. 2.2], na který případně zájemce odkazujeme. V následujících podkapitolách se podíváme na použité syntaxe zápisu jen stručně a uvedeme vysvětlující ukázkové příklady.

3.1 RDF/XML

Výchozí standard Resource Description Framework (RDF) je definován konsorciem W3C (viz <http://www.w3.org/RDF/>) a pro potřeby Dublin Core (DC) je specifikován [3] jako jednoduchý nebo kvalifikovaný zápis Dublin Core v RDF/XML. Zápis kvalifikovaného Dublin Core je v přípravě, a proto vycházíme především ze zápisu jednoduchého Dublin Core v RDF/XML.

Příklad zápisu v RDF/XML:

```
<?xml version="1.0" encoding="utf-8" ?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-  
rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/"  
xmlns:dcterms="http://purl.org/dc/terms/" xmlns:thesis="http://www.eVSKP.cz/scheme/thesis/">  
<rdf:Description>  
  <dc:type>Diplomová práce</dc:type>  
  <dc:title xml:lang="cs" >Úvod do problematiky veřejného výpůjčního práva</dc:title>  
  <dc:language>cs</dc:language>  
  <dc:creator>Kábrtová, Jana</dc:creator>  
  <dc:contributor.advisor>Příbylová, Daniela</dc:contributor.advisor>  
  <dc:contributor.referee>Kera, Denisa</dc:contributor.referee>  
  <dc:date.created>2001-10-04</dc:date.created>  
  <thesis:degree.name>Mgr.</thesis:degree.name>  
  <thesis:degree.level>magisterský</thesis:degree.level>  
  <thesis:degree.grantor>Univerzita Karlova v Praze. Filozofická fakulta. Ústav informačních  
studií a knihovnictví</thesis:degree.grantor>  
  <dc:publisher>KSV Jinonice</dc:publisher>  
  <dc:identifier>
```

```

    <dcterms:URI>http://doi.eVSKP.cz/120.12/31</dcterms:URI>
  </dc:identifier>
  <dc:description>Práce se zabývá problematikou veřejného výpůjčního práva (dále VVP) jak v
  historickém kontextu, tak jeho současnou praxí v existujících světových systémech. Práce je
  rozdělena do sedmi kapitol. -- První kapitola je věnována světové terminologii užívané v oblasti
  VVP... </dc:description>
</rdf:Description>
</rdf:RDF>

```

3.2 XML

Standard XML je pro potřeby Dublin Core podrobněji specifikován ve zvláštním doporučení [3]. Na rozdíl od implementace jednoduchého Dublin Core bude XML použité pro metadata eVŠKP obsahovat další prvky definované v návrhu metadat [5] nad rámec Dublin Core [2], což výše zmíněné doporučení také zahrnuje.

Příklad zápisu v XML:

```

<?xml version="1.0" encoding="utf-8" ?>
<metadata xmlns="http://www.eVSKP.cz/scheme/" xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:thesis="http://www.eVSKP.cz/scheme/thesis"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="
http://www.eVSKP.cz/scheme/schema.xsd">
  <dc:type>Diplomová práce</dc:type>
  <dc:title lang="cs">Úvod do problematiky veřejného výpůjčního práva</dc:title>
  <dc:language xsi:type="dcterms:RFC3066">cs</dc:language>
  <dc:creator>Kábrtová, Jana</dc:creator>
  <dc:contributor.advisor>Příbylová, Daniela</dc:contributor.advisor>
  <dc:contributor.referee>Kera, Denisa</dc:contributor.referee>
  <dc:date.created xsi:type="dcterms:w3cdtf">2001-10-04</dc:date.created>
  <thesis:degree.name>Mgr.</thesis:degree.name>
  <thesis:degree.level>magisterský</thesis:degree.level>
  <thesis:degree.grantor>Univerzita Karlova v Praze. Filozofická fakulta. Ústav informačních
  studií a knihovnictví</thesis:degree.grantor>
  <dc:publisher>KSV Jinonice</dc:publisher>
  <dc:identifier xsi:type="dcterms:URI">http://doi.eVSKP.cz/120.12/31</dc:identifier>
  <dc:description lang="cs">Práce se zabývá problematikou veřejného výpůjčního práva (dále
  VVP) jak v historickém kontextu, tak jeho současnou praxí v existujících světových systémech.
  Práce je rozdělena do sedmi kapitol. -- První kapitola je věnována světové terminologii užívané
  v oblasti VVP...</dc:description>
</metadata>

```

3.3 HTML/XHTML

Kódování v HTML/XHTML je vhodné pro účely kódování metadat do HTML/XHTML stránky s obsahem e-VŠKP.

Jednotlivé kvalifikátory jsou kódovány jako XHTML značka META,

- atribut name obsahuje jméno kvalifikátoru
- atribut content obsahuje hodnotu (obsah) kvalifikátoru
- atribut lang může obsahovat upřesnění použitého jazyka
- atribut scheme může obsahovat upřesnění použitého schématu zápisu

Příklad zápisu v HTML/XHTML:

```
...
<head profile="http://dublincore.org/documents/dcq-html/">
<title>Ukázka zápisu v HTML/XHTML</title>
<link rel="schema.DC" URL="http://purl.org/dc/elements/1.1/">
<link rel="schema.DCTERMS" URL="http://purl.org/dc/terms/">
<meta name="DC.type" content="Diplomová práce">
<meta name="DC.title" xml:lang="cs" content="Úvod do problematiky veřejného výpůjčního práva">
<meta name="DC.language" scheme="DCTERMS:RFC3066" content="cs">
<meta name="DC.creator" content="Kábrtová, Jana">
<meta name="DC.contributor.advisor" content="Příbylová, Daniela">
<meta name="DC.contributor.referee" content="Kera, Denisa">
<meta name="DC.date.created" scheme="DCTERMS:W3CDTF" content="2001-10-04">
<meta name="THESIS.degree.name" content="Mgr.">
<meta name="THESIS.degree.level" content="magisterský">
<meta name="THESIS.degree.grantor" content="Univerzita Karlova v Praze. Filozofická fakulta. Ústav informačních studií a knihovnictví">
<meta name="DC.publisher" content="KSV Jinonice">
<meta name="DC.identifier" scheme="DCTERMS:URI"
content="http://doi.eVSKP.cz/120.12/31">
<meta name="DC.description" xml:lang="cs" content="Práce se zabývá problematikou veřejného výpůjčního práva (dále VVP) jak v historickém kontextu, tak jeho současnou praxí v existujících světových systémech. Práce je rozdělena do sedmi kapitol. -- První kapitola je věnována světové terminologii užívané v oblasti VVP...">
</head>
...
```

3.4 Schémata kódování dat

Kromě tří výše zmíněných zápisů metadat obsahuje návrh standardu také specifikaci zápisu obsahu jednotlivých metadatových prvků. V aktuální verzi je definován formát zápisu pro Datum, Jazyk, Formát zdroje a Jména fyzických osob.

Datum se zapisuje podle ISO 8601, přípustné jsou různé varianty od zápisu jen roku přes zápis konkrétního dne (preferováno) až po přesnou definici času. Položky dne jsou odděleny pomlčkou, v případě času je povinný znak T a uvedení časové zóny (pro ČR „+01:00“).

Jazyk je definován dle RFC 3066, možné varianty zápisu jsou dle ISO 639-2 (kód jazyka složený ze tří znaků se specifikací země, kde je jazyk používán) nebo ISO 639-1 (kód jazyka složený ze dvou znaků).

Formát zdroje respektuje běžně používaný standard Internet Media Types - MIME (Multipurpose Internet Mail Extensions) popsany v RFC 2046.

Jména fyzických osob se zapisují ve formě: Příjmení, Křestní Jména.

Pro další typy metadat, jako jsou např. přidělované tituly, předpokládáme zřízení číselníků dostupných v rámci Národního registru metadat.

Literatura

- [1] BRATKOVÁ, E. (zprac.). *Systémy pro rozšiřování a zpřístupňování elektronických disertačních prací : analýza současného stavu v zahraničí* [online]. Praha, 2003-10-31, poslední aktualizace 2004-02-22 [2006-05-05]. 59 s. Zpráva. FRVŠ č. 2045/2003. Dostupný z WWW: <<http://www.evskp.cz/Dokumentyver/systemy-analyza.pdf>>.
- [2] *Dublin Core Metadata Element Set, Version 1.1: Reference Description* [online]. Dublin : OCLC, 2004-12-20 [cit. 2006-05-05]. Dostupný z WWW: <<http://dublincore.org/documents/2004/12/20/dces/>>.
- [3] *Dublin Core Metadata Initiative: Encoding Guidelines* [online]. Dublin: OCLC, 2003-12-09 [cit. 2006-05-05]. Dostupný z WWW: <<http://www.dublincore.org/resources/expressions/>>.
- [4] *ETD-MS : an Interoperability Metadata Standard for Electronic Theses and Dissertations* [online]. Version 1.00, revision 2. 2001-06-05, last updated on 2006-03-21 [cit. 2006-05-05]. Dostupný z WWW: <<http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html>>.
- [5] *EVSKP-MS : Metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR : Návrh* [online]. Verze 0.1. 2005-11-07 [2006-05-05]. 23 s. Dostupný z WWW: <<http://www.evskp.cz/standardy/metadata/navrh-2005-11-07.html>>.
- [6] *METADISS : Format des Metadatenatzes für Online-Hochschulschriften der Deutschen Bibliothek und des Projekts "Dissertationen Online"* [online]. Version 1.5. Stand 2003-05-05 [cit. 2006-05-05]. Dostupný z WWW: <http://deposit.ddb.de/metadiss.htm#top_I>.
- [7] World Wide Web Consortium. *RDF/XML Syntax Specification (Revised) : W3C Recommendation 10 February 2004* [online]. 2004 [cit. 2006-05-05]. Dostupný z WWW: <<http://www.w3.org/TR/rdf-syntax-grammar/>>.

PhDr. Eva Bratková

Univerzita Karlova v Praze
Ústav informačních studií a knihovnictví
U Kříže 8-10
158 00, Praha 5 Jinonice
e-mail: eva.bratkova@ff.cuni.cz

Ing. Jan Mach

Vysoká škola ekonomická
Centrum informačních a knihovnických služeb
nám. W. Churchilla 4
130 67, Praha 3
e-mail: machj@vse.cz